

NOEC AND LOEC DATA SHOULD NO LONGER BE GENERATED OR USED

Michael St J. Warne^{1*} and Rick van Dam²¹ Centre for Environmental Contaminants Research, CSIRO, PMB 2, Glen Osmond, South Australia, 5064, Australia.² Environmental Research Institute of the Supervising Scientist (ERISS), GPO Box 461, Darwin, Northern Territory, 0801, Australia.

In ecotoxicology we generate either hypothesis-based or point estimate toxicity data. The former consist of no observed effect concentrations (NOECs), lowest observed effect concentrations (LOECs) and maximum acceptable threshold concentrations (MATCs – the geometric mean of a NOEC and a LOEC). Examples of the latter consist of lethal (LC), effect (EC) or inhibition (IC) concentration data for specific effect sizes such as 50%, 20% and 10%.

The use of NOEC and LOEC data in ecotoxicology and particularly regulatory aspects of ecotoxicology has been severely criticised since a series of articles in the 1990s (e.g., Hoekstra and Van Ewijk 1993a; Noppert et al. 1994; Chapman et al. 1996; OECD 1998). Yet, despite these criticisms, NOECs and LOECs are still generated and reported regularly in the literature. For example, all of the direct toxicity assessment programs set up to examine the toxicity of the discharges from Australian desalination plants that the primary author has worked on in the last two years, have generated NOEC and LOEC data.

Given this situation, the following questions must be asked: (a) why do ecotoxicologists continue to generate NOEC and LOEC data?; (b) why do ecotoxicologists continue to use these data?; and (c) why do regulators continue to accept such data? We believe that the answers are threefold: (1) a lack of understanding of the limitations of these data; (2) a lack of understanding of the theoretical basis of the statistics used to generate these values; and (3) no organisation has clearly and categorically stated that we should no longer generate NOEC and LOEC data and, from this point on, that any new NOEC and LOEC data that are generated should not be used. In this opinion piece we briefly review the existing literature and substantiate the answers provided above. Further, we advocate that the Australasian Society for Ecotoxicology adopt a policy position that actively discourages ecotoxicologists from generating NOEC and LOEC data and prevents the publication of such data in the *Australasian Journal of Ecotoxicology* without a sound scientific justification.

The many problems with NOEC and LOEC data fall into three main categories:

- the misleading nature of their names;
- the inappropriateness of the method by which they are calculated; and
- the validity of the statistical methods used.

Misleading names

The term NOEC is misleading. It would be entirely appropriate for lay people, stakeholders and regulators who do not have extensive ecotoxicology knowledge, to assume that a NOEC is in fact a concentration that causes no toxic effect. However, we know this is not necessarily the case. By definition the NOEC is the highest concentration used in a toxicity test that does not cause a toxic effect that is significantly (usually set at $p \leq 0.05$) different to the control. This is not, by any means, the same as causing no effect. The magnitude of the biological effect that NOECs typically cause is 10 to 30% (USEPA 1991; Hoekstra and Van Ewijk 1993a; Moore and Caux 1997) and, obviously, LOECs will correspond to even larger biological effects. We do note, however, that a NOEC can be a reliable estimator of no effect when sufficient effort is placed on identifying it; for example, by the use of multiple tests with high precision, systematically homing in on the region of interest. However, this is typically not done. Also, NOECs can be more conservative than low effect concentrations, but typically only in the event of very strong threshold responses or where test precision is very high.

With the recent spate of developing Australian desalination plants and resulting studies assessing the toxicity of their discharges, some consultants and regulators have expressed concern over not using NOEC data when they were available. Although EC10 data were ultimately used rather than NOEC data (e.g., Hobbs and Warne 2008), these consultants and regulators remained concerned because (1) they felt that stakeholders would think they were not being as protective of the environment as if NOEC data were used and (2) EC10 data were not used in the Australian and New Zealand water quality guidelines (WQGs; ANZECC and ARMCANZ 2000). Recent ecotoxicological studies involving industries in northern Australia (specifically mining and mineral processing) have also resulted in low effect EC values being accepted by both industries and regulators as appropriate alternatives to NOECs (e.g., van Dam et al. 2008). However, on all occasions, the acceptance of low effect EC values occurred only after much discussion and justification in meetings and reports that would not have been necessary if there was clear national or international guidance on this issue.

It could be argued that the problem of the misleading terminology could be resolved by simply changing the term NOEC to the 'Not statistically different from the control concentration (NSDC)' or similar. However, we feel that this is only a cosmetic change and does not go anywhere near far enough in addressing the problems with NOEC and LOEC data, as outlined below.

Inappropriateness of the data to the intended use

These issues have been extensively covered in articles by Hoekstra and Van Ewijk (1993a,b), Noppert et al. (1994) and Chapman et al. (1996) and readers should refer to these articles for more detail. Here we merely summarise their comments.

NOEC and LOEC values are generally determined by using analysis of variance followed by multiple comparison tests (e.g., Dunnett's tests). Put simply, these methods use the variability in the data to determine which treatments are significantly different from the control. However, these methods can only determine which of the treatments used in a toxicity test are significantly different from the control. Given their method of calculation, NOEC and LOEC values are controlled by the concentrations of the test chemicals used in the treatments, the variability in the data, the selected significance level, and the sample size (increasing replication reduces the variability).

Critics have argued that such a method of determining NOEC values is open to abuse. Experiments conducted using poor laboratory practice will report larger variability; hence, the differences between the control and treatments will have to be larger in order to be significantly different. This has led many critics to state that 'they could design a toxicity test to give any particular NOEC value that is desired'. Regulatory authorities have attempted to limit the potential for manipulating the resulting NOEC and LOEC values by limiting maximum differences between the concentrations in the treatments (e.g., OECD (1996) recommends a maximum factor of 3.2) as well as the variability (i.e., coefficient of variation) among control replicates. However, most ecotoxicology papers do not mention the coefficient of variation of their controls and whether these meet the acceptance criteria for that particular toxicity test.

Due to the fact that NOEC and LOEC values are at least partly determined by the variability in the data, they do not correspond to any specific biological effect (as noted above they typically correspond to a 10 – 30% effect). Therefore, they are innately more variable than point estimates. This is acknowledged as the reason that relationships between toxicity and soil physicochemical properties, which are the terrestrial equivalent of the hardness correction equations in the Australian and New Zealand water quality guidelines (ANZECC and ARMCANZ 2000), based on NOECs generally have lower coefficients of determination (r^2) than relationships based on point estimates (Rooney et al. 2006).

This ambiguity in the level of protection provided by a NOEC flows through to environmental quality guidelines derived using such data. For example, a PC95 based on chronic sub-lethal NOEC data should theoretically protect 95% of species from experiencing statistically significant chronic sub-lethal toxic effects. In contrast, a PC95 based on chronic sub-lethal EC10 data should theoretically protect 95% of species from experiencing chronic sub-lethal effects greater than 10% larger than the modelled or mean control response.

The usual threshold of probability used to determine statistical significance is 0.05. However, Fisher, the originator of this threshold 'did not advocate dogmatic application of any particular threshold p ' (Newman 2008). By adopting a threshold of 0.05 it means that we are only prepared for there to be a 5% chance that we are wrong when we say there is a significant difference (i.e., 5% probability of a Type I error). But what is more appropriate for us as ecotoxicologists (if we persist with the concept of NOECs) is what is the probability that we are wrong when we say that there is no significant difference (i.e., the probability of making a Type II error, also known as β)? This concept was first introduced by Neyman and Pearson and is philosophically different to Fisher's approach (see Newman 2008 for details). In ecotoxicology, we should want a similarly low probability (e.g., 5%) of being wrong when we state that a chemical concentration is not significantly different from the control.

Greater attention needs to be paid to the probability of committing a Type II error and, related to this, the statistical power of a test. The power of a test represents the probability of correctly rejecting the null hypothesis, and is defined as $1 - \beta$. Thus, as power increases, the probability of committing a Type II error decreases. Power analysis should typically be done before the test (i.e. *a priori*) to determine the most appropriate experimental design (e.g., sample size, minimum detectable difference) for a test with appropriate pre-specified Type I and Type II error rates. However, unfortunately, power analysis is seldom reported in ecotoxicology. An example of the implications of a lack of power is provided by Broos et al. (2007) who showed that 9, 10, 19, 27, 36 and 93 replicates were required in order to detect a 20% difference in soil microbial biomass carbon (at $p = 0.05$) in six different field trials. Considering that the usual degree of replication for such experiments is 3 or at best 5, then the associated ANOVA and multiple comparison tests would not be able to detect that a 20% effect was significantly different.

Alternatively, confirmation of statistical power can be obtained *post hoc* by determining the minimum significant difference (MSD). The MSD is the minimum difference (usually stated as a percentage) between the mean values of the control and a treatment that can be determined to be significantly different given the experimental design used and data. The MSD increases as the variation within treatments and controls increases. Therefore, if MSDs are presented along with NOEC and LOEC values then users will be provided with an idea of the variability of the test that generated the NOEC and LOEC values and how reliable they are. A workshop held to discuss the appropriateness of using NOEC and potential replacement measures in fact recommended that if NOEC data are still to be generated that they should always report the MSD and the difference between the NOEC and the control (i.e., the magnitude of the biological effect caused by the NOEC) (Van der Hoeven 1997). It should, however, be noted that *post hoc* analysis of power and calculation of MSDs has been criticised as being uninformative (Newman 2008). One of the most common statistical packages used in ecotoxicology, ToxCalc™, provides MSD values in its standard toxicity test summary reports.

Validity of the statistical methods used to determine NOEC and LOEC data

Newman (2008) has written a highly critical article on the methods used to derive NOEC and LOEC data which should be read by all ecotoxicologists. In the article he also argues that these methods are fundamentally flawed. It is clear from the literature he cites that this issue has been raised in many other branches of science including psychology, ecology, sports medicine, medicine, forestry and conservation biology. The flaws are so large that many medical journals (e.g., *The Lancet*, the *British Medical Journal*, the *Medical Journal of Australia*, the *American Journal of Public Health* and the *British Heart Journal* (Altman et al. 2000)), the International Committee of Medical Journal Editors (ICMJE 1988) and the American Psychological Association Publication Manual (APA 2001; which sets editorial standards for more than 1000 journals) have developed guidelines which strongly recommend that hypothesis-based tests of significance are not to be used or accepted for publication and that they be replaced by the use of confidence intervals. Ultimately, Newman (2008) recommends that statistical significance methods should be replaced 'whenever possible' by confidence interval-based methods.

What are the alternatives to NOEC and LOEC data?

There are several alternatives to NOEC and LOEC data including: low percent effect point estimates (e.g., 5, 10 and even 20% levels) termed the ECx approach; a parametric estimate of the no effect concentration (NEC; Van der Hoeven 1997); and a bounded effect concentration (BEC; Hoekstra and Van Ewijk 1993b). Key hindrances to the acceptance and adoption of the NEC approach are doubts over whether a threshold concentration exists and the concern that there will seldom be sufficient data available to validate the choice of model used to determine the NECs (Van der Hoeven 1997). Fox (in press) has proposed the use of NEC estimation using an alternative Bayesian approach. This proposal may have merit, but requires additional testing and comparison with other methods before being considered for adoption.

A BEC can be calculated in two ways. The simplest way is to determine the highest tested concentration that has an upper 95% confidence interval (CI) that causes less than the selected % effect (e.g., a BEC10 is the highest tested concentration that has an upper 95% CI that causes less than a 10% effect). The more complex way is to determine the highest tested concentration which rejects the hypothesis 'the toxic effect is y%' where y is typically 25 (Hoekstra and Van Ewijk 1993b). Using linear regression from this point to zero, the concentration that causes the desired effect is determined (e.g., extrapolated to the 10% effect which becomes the BEC10). While the BEC has merit, it tends to generate conservative numbers and is not intuitively straightforward and, perhaps as a consequence, has not been widely adopted with a few exceptions (Markich and Camilleri 1997; Franklin et al. 2000). Thus, the tendency has been to replace NOEC and LOEC data with low percent effect point estimates.

Low percent effect point estimate data are not without their own limitations (see also the concerns outlined by Fox (2008)). The principal limitation of low effect data is that, as the percent effect decreases, the error associated with the estimate increases, all other things being equal. Therefore, there is greater uncertainty in what the low percent effect value actually is. In addition to this uncertainty, estimates of effects below the 10% level are typically more model dependent than effects closer to the median level, particularly when the estimate is based on extrapolation outside the data range (Moore and Caux 1997). However, uncertainty can be greatly reduced if the estimate of toxicity is determined by interpolating between actual data rather than extrapolating from data (Stephan and Rogers 1985; Moore and Caux 1997). Nonetheless, when point estimate toxicity data are used the variability associated with their use is made clear to the reader, whereas the use of NOEC and LOEC data (which have no associated measure of variability) give the unwary reader a completely false impression of their variability, unless the MSD is reported - which invariably is not the case. Another limitation is that with the standard toxicity test designs there are usually insufficient data to determine the best fitting model (Van der Hoeven 1997) from a group of biologically plausible models. Perhaps the key hindrance to replacing NOEC and LOEC data with ECx data is that, despite the body of work and level of debate throughout the 1990s, there has been no definitive statement by any regulatory or standardisation authority as to which percent effect should be reported in toxicity tests. Van der Hoeven (1997) states, correctly, that there is no objective method to select the most appropriate ECx value. However they do provide some guidance - that the:

- “(i) x should be small because an (almost) no effect level is intended;
- (ii) x should not be too small because of problems of accuracy and model dependence; and
- (iii) x should be a round number”.

Van der Hoeven (1997) concluded that a EC5 or EC10 should be used. Our observation from the literature is that when low effect EC values have been reported it is typically the EC10.

Warne (1998), when developing the framework for deriving water quality Trigger Values (TVs) for toxicants that was adopted in the current Australian and New Zealand WQGs (ANZECC and ARMCANZ 2000), recommended that NOEC data be used to derive the WQGs “but that their use be phased out as LC5 type data become available”. Going a step further CCME (2007) recently specified a preference for EC10 data over a number of other toxicity estimates, including NOECs, for use in the derivation of water quality guidelines for toxicants in Canada.

Another factor slowing adoption of the ECx approach is that, in order to obtain better low percent effect point estimates, it will be necessary to modify our current experimental designs. More low concentrations will need to be tested and concentrations that cause high percent effects (e.g., 50% or

higher) will be less important than presently is considered to be the case. In addition, estimating low percent effects is best done using regression analysis and for such analysis increasing the number of treatments is more beneficial than having replicates (Stephan and Rogers 1985; Moore and Caux 1997). However, such changes can easily be incorporated into experimental designs.

Where to from here?

Some (de Bruijn and Hof, 1997) have questioned “whether it is worth embarking on this process” of replacing NOEC and LOEC data. We do not support this argument. We believe that, based on the reality that what is currently being done is wrong, corrective action is required. Earlier in this opinion piece we argued that we need to ask: (a) why ecotoxicologists continue to generate NOEC and LOEC data?; (b) why ecotoxicologists continue to use these data?; and (c) why do regulators continue to accept such data? We hope that this opinion piece has explained why NOEC and LOEC data should no longer be generated.

We believe the answer to the third question we posed is that no organisation has clearly and categorically stated that we should no longer generate, publish and use NOEC and LOEC data. Although an OECD (1998) workshop recommended that NOECs be phased out of international standards, guidance on hypothesis testing, NOEC and LOEC estimations remain in more recent OECD statistical publications (i.e. OECD 2006) because “...the NOEC is still required in many regulatory standards from many countries and in some cases where a detailed determination of an EC_x is not relevant and the alteration of the study design is too costly to fulfil the requirements for regression models”. Environment Canada has addressed the issue to some extent in that they now: (i) no longer include the methods for calculating NOEC and LOEC values in their toxicity testing protocols (R Scroggins, *pers. comm.* 2008); and (ii) have a stated preference for EC₁₀ data over NOEC data for the derivation of water quality guidelines (CCME 2007). However, they also have not come out and banned their calculation or use.

Australia has a proud history of being innovators in the area of ecotoxicology. Let us continue this tradition, let us bite the bullet and as of this date no longer generate or use NOEC and LOEC data. We urge the Australasian Society for Ecotoxicology to adopt a policy position that actively discourages ecotoxicologists from generating NOEC and LOEC data and prevents the publication of such data in the *Australasian Journal of Ecotoxicology* without a sound scientific justification (e.g. to enable direct comparison to previous studies). We believe that NOEC and LOEC data generated after a certain date should no longer be used or accepted for publication because without adopting such a position there is little incentive for scientists to make the switch. The likelihood of not having your data being published or used in Australia would be a major incentive to ‘do the right thing’.

However, in adopting such a position, we realise a reasonable and robust alternative must be offered and, most importantly, accepted broadly by researchers, regulators and industry.

Some alternatives have been introduced very briefly in this paper, and Fox (2008) also addresses this. At present, low effect EC values (most likely EC₅s or EC₁₀s) have the widest acceptability as alternatives to NOECs, and recent developments in water quality guideline derivation processes mean that they are now likely to be considered acceptable and used for guideline derivation purposes. Consequently, we feel that low effect EC values should be reported (with accompanying confidence limits) instead of NOECs. The likely revision of the Australian and New Zealand water quality guidelines (ANZECC and ARMCANZ 2000) over the next two or three years presents the perfect opportunity for Australia to take the lead on this statistical issue and progress it to the point where unambiguous and sound guidance can be provided to correct a long recognized ecotoxicological data analysis error.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the useful comments on the manuscript by Dr Rick Krassoi and Prof. Richard Lim and two anonymous reviewers.

REFERENCES

- Altman DG, Machin D, Bryant TN and Gardner MJ. 2000. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. 2nd Ed. British Medical Journal Books, London, England.
- APA (American Psychological Association). 2001. *Publication Manual of the American Psychological Association*. 5th Ed. American Psychological Association, Washington, DC, USA.
- ANZECC and ARMCANZ (Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand). 2000. *National Water Quality Management Strategy, Australian and New Zealand Guidelines for Fresh and Marine Water Quality*. ANZECC and ARMCANZ, Canberra, Australia.
- Broos K, Macdonald LM, Warne MStJ, Heemsbergen DA, Barnes M, Bell M and McLaughlin MJ. 2007. Limitations of soil microbial biomass carbon as an indicator of soil pollution. *Journal of Soil Biology and Biochemistry* **39**, 2693-2695.
- CCME (Canadian Council of Ministers of the Environment). 2007. A protocol for the derivation of water quality guidelines for the protection of aquatic life 2007. In: *Canadian Environmental Quality Guidelines*, Canadian Council of Ministers of the Environment, 1999, Winnipeg, MB, Canada. 37 pp.
- Chapman PM, Cardwell RS and Chapman PF. 1996. A warning: NOECs are inappropriate for regulatory use. *Environmental Toxicology and Chemistry* **15**, 77-79.
- De Bruijn JHM and Hof M. 1997. How to measure no effect. Part IV: How acceptable is the EC_x from an environmental policy point of view? *Environmetrics* **8**, 263-267.
- Fox DR. in press. An integrated Bayesian approach for determining the no effect concentration and hazardous

- concentration in ecotoxicology. *Ecotoxicology and Environmental Safety*.
- Fox DR. 2009: NECs, NOECs and the ECx. *Australasian Journal of Ecotoxicology* **14**, 7-10.
- Franklin N, Stauber J, Markich S and Lim R. 2000. pH-dependent toxicity of copper and uranium to a tropical freshwater alga (*Chlorella* sp.). *Aquatic Toxicology* **48**, 275-289.
- Hobbs D and Warne MStJ. 2008. *Toxicity Assessment for the Victorian Desalination Project*. Report submitted to the Victorian Department of Sustainability and Environment. 432 pp.
- Hoekstra JA and Van Ewijk PH. 1993a. Alternatives for the no-observed effect level. *Environmental Toxicology and Chemistry* **12**, 187-194.
- Hoekstra JA and Van Ewijk PH. 1993b. The bounded effect concentration as an alternative to the NOEC. *Science of the Total Environment* **134**, Supplement 1, 705-711.
- ICMJE (International Committee of Medical Journal Editors). 1988. Uniform requirements for manuscripts submitted to biomedical journals. *Annals of Internal Medicine* **108**, 258-265.
- Markich S and Camilleri C. 1997. *Investigation of Metal Toxicity to Tropical Biota: Recommendations for Revision of the Australian Water Quality Guidelines*. Supervising Scientist Report 127, Supervising Scientist, Canberra, ACT, Australia.
- Moore DRJ and Caux PY. 1997. Estimating low toxic effects. *Environmental Toxicology and Chemistry* **16**, 794-801.
- Newman MC. 2008. What exactly are you inferring? A closer look at hypothesis testing. *Environmental Toxicology and Chemistry* **27**, 1013-1019.
- Noppert F, Van der Hoeven N and Leopold A. 1994. *How to Measure No Effect? Towards a New Measure of Chronic Toxicity in Ecotoxicology*. Netherlands Working Group on Statistics and Ecotoxicology, Delft, The Netherlands.
- OECD (Organisation for Economic Co-operation and Development). 1996. *Daphnia magna reproduction test*. Proposal for Updating Guideline 202, Part II, Revised Draft Document, Paris, France. 22 pp.
- OECD. 1998. *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data*. OECD Environmental Health and Safety Publication, Series on Testing and Assessment, No. 10, Environment Directorate, ENV/MC/CHEM(98)18, Paris, France. 133 pp.
- OECD. 2006. *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application*. OECD Series on Testing and Assessment No. 54, Environment Directorate, ENV/JM/MONO(2006)18, Paris, France. 147 pp.
- Rooney C, Zhao FJ and McGrath SP. 2006. Soil factors controlling the expression of copper toxicity to plants in a wide range of European soils. *Environmental Toxicology and Chemistry* **25**, 726-732.
- Stephan CE and Rogers JW. 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. In *Aquatic Toxicology and Hazard Assessment*. STP 891. Eds RC Bahner and DJ Hansen, American Society for Testing and Materials, Philadelphia, PA, USA, pp 328-338.
- USEPA (United States Environmental Protection Agency). 1991. Technical Support Document for Water Quality-Based Toxics Control. Report No. EPA 505/2-90-001. Washington, DC, USA.
- van Dam RA, Hogan AC, Harford AJ and Markich SJ. 2008. Toxicity and metal speciation characterisation of waste water from an abandoned gold mine in tropical northern Australia. *Chemosphere* **73**, 305-313.
- Van der Hoeven N. 1997. How to measure no effect. Part 1: Towards a new measure of chronic toxicity? *Environmetrics* **8**, 241-248.
- Warne MStJ. 1998. Critical Review of Methods to Derive Water Quality Guidelines for Toxicants and a Proposal for a New Framework. Supervising Scientist Report 135, Supervising Scientist, Canberra, ACT, Australia. ISBN 0 642 24338 7. 82 pp.